安全科学与工程学报
Journal of Safety Science and Engineering
第 4 卷◆第 3 期◆版本 1.0◆2026 年
文章类型：论文丨刊号（ISSN）：2972-4716(P) / 2972-4724(O)

“人工智能+”技术在信息安全领域中的融合应用

梁钰
广西开放大学
DOI:10.32629/jsse.v4i3.20986

[摘　要] 人工智能技术的持续革新与落地应用,为信息安全行业的转型升级提供了全新的技术路径与核心动能。本文立足行业发展实际,围绕威胁检测预警、漏洞挖掘修复、身份认证管控、攻击溯源应急四大核心应用场景,客观研判二者融合发展过程中存在的技术安全隐患、产业落地壁垒、人才资源短缺、标准体系缺失等现实问题,提出适配行业发展的优化路径,为网络空间安全建设提供理论与实践参考。
[关键词] 人工智能+；信息安全；网络攻击；安全防护
中图分类号：TP18　文献标识码：A

Integrated Application of "AI +" Technology in the Field of Information Security

Yu Liang
Guangxi Open University

[Abstract] The continuous innovation and practical deployment of artificial intelligence technology have provided brand-new technical approaches and core driving forces for the transformation and upgrading of the information security industry. Based on the actual development of the industry, this paper focuses on four core application scenarios: threat detection and early warning, vulnerability mining and remediation, identity authentication and management, as well as attack traceability and emergency response. It objectively analyzes practical problems arising from the integrated development of the two, including potential technical security risks, industrial implementation barriers, shortage of professional talents and incomplete standard systems. Corresponding optimization strategies tailored to industrial development are proposed, aiming to offer theoretical and practical references for the construction of cyberspace security.
[Key words] AI +; Information Security; Cyber Attack; Security Protection

1 引言

在计算机网络技术迅速发展的带动下,人工智能与时俱进,直接间接地在推动其他领域的进步,网络安全问题日渐成为人们关注的焦点。人工智能在信息安全领域中得到了广泛应用。技术革新的同时,网络安全形势持续恶化,网络攻击告别传统零散模式,呈现出智能化、自动化、规模化的特征。APT攻击、勒索软件、DDoS攻击及大规模数据泄露事件频繁发生,不仅威胁个人信息与企业经营安全,更严重危害金融、能源、政务等关键信息基础设施的稳定运行,已然成为制约数字经济健康发展的重要因素。现阶段,“人工智能+”与信息安全的融合创新持续深化,逐步重塑网络安全的防护架构与运营模式,成为行业发展的主流趋势。

需要正视的是,人工智能与信息安全的融合具有双面性,在赋能安全防护的同时,也催生了全新的网络安全风险。一方面,攻击者可借助人工智能优化攻击手段,批量生成恶意代码、制作对抗样本、实施精准钓鱼攻击,大幅降低攻击门槛,提升网络攻击的隐蔽性与破坏力；另一方面,人工智能模型存在可解释性差、高度依赖优质训练数据、算法偏见等固有缺陷,制约了智能安全技术的落地应用。在此背景下,系统研究人工智能技术在信息安全领域的应用场景,梳理当前融合发展的现实难题,制定针对性的优化发展策略具有重要的理论价值与现实意义。

2 “人工智能+”技术在信息安全领域中融合的核心应用场景

2.1威胁检测与预警

威胁检测与预警是网络安全主动防御的核心前置模块,也是人工智能在信息安全领域落地最成熟的应用场景。人工智能技术通过对网络常态行为的建模分析,构建动态运行基线,可精准识别各类异常行为,实现未知威胁的主动预判。在网络流量监测中,深度学习模型能够挖掘流量时序与统计特征,有效识别DDoS攻击、代码注入、跨站脚本等常见网络攻击,提升复杂网络

环境的检测精度。在终端防护层面，人工智能可实时采集终端进程、文件、网络连接等运行数据，动态甄别恶意程序与异常操作，同步完成恶意代码的分类研判。针对内部安全风险，基于人工智能的用户行为分析系统可积累用户日常操作、登录、资源访问等行为数据，构建个性化行为画像，及时发现账号盗用、越权访问等违规行为[1]。

2.2漏洞挖掘与修复

系统漏洞是网络入侵的主要突破口，漏洞挖掘与修复是筑牢网络安全防线的关键环节。传统漏洞检测以静态分析、动态测试和随机模糊测试为主，高度依赖人工经验，不仅检测效率偏低、误报漏报率较高，且难以识别代码深层的隐性逻辑漏洞。人工智能技术的应用，推动漏洞挖掘从人工经验驱动转向智能数据驱动，大幅提升漏洞排查的效率与精准度。在静态检测场景下，机器学习与自然语言处理技术可深度解析源代码与二进制代码，提取代码语法与语义特征，精准识别常规漏洞与传统技术难以发现的隐性漏洞。在动态测试环节，强化学习优化的智能模糊测试技术可自主学习程序运行路径与历史漏洞特征，定向生成高质量测试用例，优先覆盖高风险代码区域，有效提升代码覆盖率与漏洞触率[2]。

2.3身份认证与访问控制

身份认证与访问控制是信息系统边界安全管控的核心屏障，直接决定网络资源的访问安全与管理规范度。传统身份认证多采用静态密码、物理令牌、单一生物识别等方式，普遍存在密码泄露、设备丢失、特征伪造等安全隐患；配套的静态权限管控模式固化僵化，无法适配动态多变的业务场景与网络环境，难以抵御复杂的越权访问与非法入侵风险。人工智能技术重构了身份认证与权限管控体系，实现了安全性与实用性的平衡。在身份认证方面，智能多因素认证体系融合人脸、指纹等生物特征与用户操作行为特征，依托算法完成多维度数据融合研判，凭借行为特征唯一性、不可复制性的优势，有效防范身份伪造、账号盗用等风险，提升认证精准度。在访问控制方面，人工智能构建动态自适应管控机制，结合用户身份、操作行为、访问场景、网络环境等多维数据，实时动态调整访问权限。当系统检测到高危访问环境或异常操作行为时，可自动限制、冻结权限或启动二次认证，从源头规避非法访问与数据泄露风险。

2.4攻击溯源与应急响应

攻击溯源与应急响应是网络安全防御的后置关键环节，承担着梳理攻击链路、追溯攻击源头、止损修复系统的重要作用。传统溯源与应急处置高度依赖人工运维研判，需要人工梳理海量日志、拼凑攻击链路，不仅效率低下、人力成本较高，且面对规模化、复杂化的新型网络攻击，极易出现溯源不全、研判不准等问题，难以适配当下网络安全实战处置需求。人工智能技术的赋能，推动安全溯源与应急处置走向自动化、智能化、精准化。在攻击溯源层面，智能分析系统可全域归集网络流量、系统运行、应用操作等全量日志数据，依托图神经网络与关联分析技术，可视化还原攻击者的入侵路径、传播方式与攻击手段，精准定位攻击源头。同时，系统可基于海量攻击数据提炼攻击者行为特征，构建攻击者画像，为后续风险防控与责任追溯提供依据。在应急响应层面，人工智能驱动的SOAR平台整合各类安全工具与运维流程，实现安全告警的自动归集、分级、研判与处置，可自主完成恶意IP拦截、感染主机隔离、恶意文件清除等基础操作。依托持续对抗学习能力，平台可动态优化处置策略，适配不同类型、不同等级的安全事件，有效解决了传统人工响应滞后、处置不规范、应对大规模攻击能力不足的问题，显著提升网络安全运营的整体效能[3]。

3 “人工智能+”赋能信息安全面临的现实挑战

3.1人工智能自身的安全风险

人工智能在赋能信息安全防护的同时，自身的技术缺陷与对抗风险也成为二者深度融合的首要阻碍。当前智能安全模型普遍存在抗干扰能力弱的问题，极易遭受对抗样本与数据投毒攻击，攻击者可通过细微的数据扰动误导模型判别结果，或在训练数据中植入恶意样本，破坏模型训练逻辑，导致安全检测出现误判、漏判，大幅削弱防护有效性。同时，深度学习模型固有的“黑箱”特性使得算法决策可解释性不足，安全人员无法溯源模型告警依据，难以核验风险真实性，既增加了安全运营压力，也降低了智能防护体系的可信度。更为关键的是，人工智能技术的开放性使其成为双向攻防工具，攻击者可依托人工智能自动化生成恶意代码、开展批量漏洞探测与精准钓鱼攻击，降低网络攻击门槛，形成智能化攻防对抗的新型安全威胁。

3.2技术落地与产业化难题

除技术本体风险外，人工智能安全技术在落地应用与产业化发展中仍存在诸多现实壁垒。高质量标注数据是模型训练的核心基础，但网络安全数据具备高度保密性与敏感性，行业数据开放共享程度低，且数据标注对专业能力要求高、成本高昂，导致多数智能安全模型训练样本不足，泛化能力薄弱，难以适配复杂真实的网络环境。同时，实验室环境下的优质模型性能，在充满噪声与未知干扰的实际场景中会明显下降，检测精度和稳定性不足，且模型训练、推理所需的算力资源，对中小微企业形成较高的成本门槛。此外，现有智能安全产品与传统防火墙、入侵检测等存量安全设备兼容性不足，系统集成与联动响应难度大，且行业缺乏统一的技术接口与落地标准，不同厂商产品难以互联互通。加之智能安全产品效果难以量化、研发投入高、回报周期长，尚未形成成熟的商业化运营模式，严重制约了产业规模化发展与中小企业创新参与[4]。

3.3人才短缺与制度滞后

复合型人才稀缺与行业制度体系滞后，是制约人工智能与信息安全融合发展的核心软性瓶颈。人工智能安全属于人工智能与网络安全交叉领域，对从业人员综合专业素养要求极高，目前国内同时精通算法模型与攻防技术的高端复合型人才缺口巨大，难以满足行业快速发展的人才需求。制度层面，我国尚未出台针对人工智能安全的专项法律法规与行业标准，在产品安全评估、模型合规性、算法透明度、智能攻击追责等领域缺乏明

确规范与监管依据,行业发展缺乏统一指引。与此同时,尽管《数据安全法》《个人信息保护法》已正式落地,但适配人工智能模型训练的数据共享、隐私保护细则仍不完善,如何平衡数据利用与隐私安全,成为制约智能安全技术迭代与数据资源复用的难题。

4 推动“人工智能+”与信息安全深度融合的发展路径

4.1加强核心技术攻关,提升自主创新能力

核心技术自主创新是人工智能与信息安全深度融合发展的核心支撑。为突破当前技术瓶颈,国家需强化人工智能安全领域的基础研究与前沿技术布局,依托重大科技专项推动校企、院所、企业协同攻关,重点突破对抗样本防御、数据投毒防护、模型可解释性优化等关键技术,同时依托联邦学习、差分隐私技术破解数据孤岛与隐私保护难题。行业企业也应持续加大研发投入,搭建专业化创新平台,深化产学研用融合模式,加速前沿技术落地转化,研发具备自主知识产权的智能安全产品,全面提升我国网络安全领域的核心技术竞争力。

4.2构建协同创新生态,促进产业健康发展

完善的产业创新生态是人工智能安全产业稳健发展的重要保障。需构建政府引导、企业主导、多方协同的创新发展体系,政府通过完善产业政策、搭建产业平台、布局产业园区,引导产业链上下游集聚发展,依托龙头企业带动中小微企业协同创新,形成融通发展的产业格局。同时,需建立标准化的网络安全数据与威胁情报共享机制,在严守数据安全与个人隐私底线的前提下,打通政企、校企的数据壁垒,实现优质安全数据资源的高效复用,为智能安全模型迭代优化与全域网络安全防护提供资源支撑。

4.3加快人才培养步伐,完善人才培养体系

复合型人才缺口是制约行业发展的关键短板,构建系统化人才培养体系是破解困境的根本举措。高校需优化学科布局,增设人工智能安全交叉学科与专业课程,定向培养兼具人工智能算法能力与网络安全专业素养的复合型人才。同时深化校企合作,搭建实训平台、推行订单式培养,提升人才实操能力。针对行业在职人员,开展常态化跨界技能培训,补齐从业人员技术短板,并通过人才引进政策吸纳海外高端人才。配套完善人才评价、激励与流动机制,打造留才、育才、引才的良好行业环境,夯实产业发展的人才根基。

4.4完善法律法规制度,加强安全监管

健全的制度规范与监管体系是人工智能安全行业规范化发展的底线保障。相关部门需加快出台人工智能安全专项法律法规与行业标准,完善智能安全产品的研发、运维、应用规范,建立统一的产品安全评估与认证体系,规范行业发展秩序。同时构建常态化、专业化的智能安全监管机制,明确监管权责,严厉打击利用人工智能实施网络攻击、数据窃取、隐私侵权等违法违规行为。此外,主动参与全球人工智能安全治理,积极对接国际行业标准,深化跨国技术与治理合作,共同应对智能化网络安全的全球性风险。

5 结论

伴随网络攻击向智能化、复杂化演变,传统规则化、被动式的信息安全防护模式已难以适配当下网络安全防护需求。将人工智能技术与信息安全领域深度融合,是网络安全体系迭代升级的必然趋势。结合当前落地现状来看,人工智能赋能网络安全仍面临多重发展瓶颈。未来,人工智能技术与网络安全产业的融合程度将持续加深。针对现存发展痛点,行业需聚焦核心技术攻关,完善产业协同生态,搭建复合型人才培养体系,健全行业监管与法律制度。通过多维度协同优化,持续释放人工智能的技术赋能价值,推动智能安全产业规范化、高质量发展。

[参考文献]

[1]辛欣.人工智能技术在信息安全管理中的应用研究[J].中国科技投资,2025,(14):137-139.

[2]李华,张远夏,杨玉.人工智能技术在信息安全中的应用[J].现代工业经济和信息化,2024,14(4):108-110.

[3]刘骁,张鹏.人工智能技术在信息安全领域中的应用研究[J].中国新通信,2022,24(17):116-118.

[4]冯抒.人工智能技术在信息安全中的应用[J].电子技术,2022,51(5):134-135.

作者简介：

梁钰(1981--),女,壮族,广西南宁人,硕士研究生,工作单位：广西开放大学,研究方向：计算机网络、信息安全。